



LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

Modelo predictivo para la selección de docentes a través de un aprendizaje supervisado

Predictive model for the selection of teachers through supervised learning

Lady Nataly Romaña Peralta

Profesional en Negocios Internacionales

lnromanap@unilibertadores.edu.co

Fundación Universitaria Los Libertadores

Director:

M Sc. Mario Gregorio Saavedra Rodríguez

mgsaavedrar@libertadores.edu.co

RESUMEN

En el presente trabajo se explora el uso del clasificador Random Forest para generar un modelo de selección de los candidatos que finalizarán el programa ofrecido por Enseña por Colombia. Random forest es un clasificador estadístico nuevo y poderoso que está bien establecido en otras disciplinas, pero es relativamente desconocido en la selección docente. Dentro de las principales ventajas se encuentran (1) un método novedoso para determinar la importancia de las variables; (2) capacidad para modelar interacciones complejas entre variables de respuestas; (3) flexibilidad para realizar varios tipos de análisis de datos estadísticos, que incluyen regresión, clasificación, análisis no



supervisados y supervisados (Cutler, 2007). En esta investigación se utiliza las variables disponibles en el proceso de selección docentes de la organización en la fase de entrevista. Se utiliza la base de datos existente desde 2015 al 2018 de los participantes seleccionados.

Palabras clave: *Random forest 1, Machinelearning 2, Teacher selection 3.*

Abstract

In this paper, the use of the Random Forest classifier is explored to generate a selection model for candidates who will complete the program offered by Enseña por Colombia. Random Forest is a powerful statistical classifier that is well established in other disciplines, but is relatively unknown in teacher selection. The main advantages of Random Forest algorithm are (1) a novel method to determine the importance of the variables; (2) ability to model complex interactions between predictor variables; (3) flexibility to perform various types of statistical data analysis, including regression, classification, unsupervised and supervised analysis(Cutler, 2007). In this research, the variables available in selecting teacher's process for the interview phase of are used. The existing database from 2015 to 2018 of the selected participants is used.

Keywords: *Random forest 1, Machine learning 2, proceso de selección docente 3.*

INTRODUCCIÓN

La sociedad civil ha buscado la forma de contribuir y cerrar las brechas de inequidad alrededor del mundo. Un claro ejemplo de ello es Teach for all, una gran red global de organizaciones independientes que tienen como propósito generar un liderazgo colectivo para asegurar que los niños, niñas y jóvenes puedan desarrollar su máximo potencial (Teach for all, 2017). Esta es una apuesta puntal hacia el sector



educativo que nace en el 2007 y su modelo alrededor del mundo es seleccionar profesionales graduados con altos estándares académicos y sociales para enseñar en colegios de altas necesidades durante dos años. Esta experiencia de inmersión permitirá a los participantes evidenciar las injusticias sociales y económicas que se viven en cada comunidad y que sean estas las que los motiven de por vida a trabajar por la equidad y justicia en cada una de las naciones. Durante los trece años de Teach for All, se han sumado 54 países a esta red, entre ellos Colombia.

Enseña por Colombia, en adelante Enseña por Colombia, ingresa al movimiento en el 2010 inquietado por la falta de oportunidades educativas para todas las personas. En el 2012 selecciona la primera cohorte de profesionales para ser docentes en la ciudad de Bogotá. Ocho años más tarde, Enseña por Colombia ha logrado llegar a estudiantes de los departamentos de Amazonas, Antioquia, Atlántico, Bolívar, Bogotá, Magdalena, Guajira y Nariño. Año a año se buscan profesionales comprometidos con el propósito que tiene la organización cuyo actuar futuro este en concordancia con las reflexiones e investigaciones que se desarrollaron en la experiencia vivida en cada uno de los contextos en los que el participante es ubicado.

Para los profesionales que llegan a ser parte de las cohortes anuales se ha desarrollado todo un programa de justicia social, liderazgo y pedagogía, que los acompaña en su experiencia y reflexiones sobre los contextos. Durante los dos años de trabajo en las comunidades, los participantes crecen en las competencias de trabajo con otros, habilidades para la enseñanza, gestión y pensamiento estratégico y mentalidad de bienestar colectivo. Todos estos conocimientos se ponen al servicio de la sociedad una vez los participantes, ahora Alumnis¹, finalizan el programa y son parte de nuevos

¹ Alumnis: participantes que terminan los dos años de programa de Enseña por Colombia



proyectos. Para el año 2019 se tenían 213 Alumnis, impactando en los siguientes sectores: 3,57% estaban estudiando, 7,74% ejercen liderazgo educativo, 8,93% Sector público, 10,12% Sector privado, 17,26% ONG o Emprendimiento, 26,79% Liderazgo escolar, 25,59% otros (Enseña, 2019).

Bajo el modelo de Enseña por Colombia y Teach for All uno de los propósitos es “desarrollar un liderazgo colectivo para asegurar que todos los niños tengan la oportunidad de desarrollar su máximo potencial”, por lo tanto, el valor humano y social de cada una de las personas que son seleccionadas es enorme, pues se tienen altas expectativas sobre su labor futura. Sin embargo, entre el 2012 y 2018 fueron seleccionados 502 profesionales para este proyecto y 272 son Alumnis, lo que quiere decir que solo 54% de los seleccionados logra finalizar el programa de dos años. Por otro lado, algunos de los efectos negativos de la deserción de los participantes son los siguientes:

- En primer lugar, detrás de cada selección de participantes se encuentran también comprometidos los intereses de otras organizaciones que como donantes y aliados han aportado a esta propuesta. Las relaciones pueden peligrar si incrementa el número de desertores.
- Los efectos en la educación de los estudiantes pueden ser nefastos, dejando estudiantes sin el docente de la materia, que en muchos casos no logra ser remplazado inmediatamente generando así inconformidades y brechas en la educación.
- En el ámbito regional educativo hay escases de algunos perfiles profesionales relacionados con las áreas de inglés, ciencias, tecnología, matemáticas y artes, por lo tanto, reclutar y seleccionar de nuevo un talento con los estándares de



calidad que ofrece Enseña por Colombia puede ser una tarea compleja para los colegios.

- Desde el costo económico que genera la deserción, de acuerdo con los datos de Enseña por Colombia 2019 la formación del instituto² de ese año fue de \$6.130.000 pesos colombianos por cada participante, una alta deserción además de las brechas sociales que genera, presenta pérdidas de recursos económicos.

Detectando esta problemática que tiene la organización, el objetivo de la investigación es aprender de las experiencias pasadas de selección, y aportar al área de reclutamiento, selección y matriculación con un modelo predictivo, random forest, basado en las evidenciadas recolectadas en el día de entrevista de los individuos para predecir quienes terminarían el programa de dos años. Cabe aclarar que en ningún momento esta investigación pretende cuestionar el modelo de selección utilizado o tener en cuenta las variables que se generan durante los dos años del programa.

REFERENTES TEORICOS

Tipos de selección de personal

Encontrar los profesionales o personas con las capacidades necesarias para ejercer determinada labor, necesita de un trabajo consiente por parte de la empresa, negocio, organización o institución que desea contratar. Esto implica reconocer que existen diversos tipos de selección de personal, que se acomodan a la necesidad que se presenta y con lo que se busca la mejor solución dependiendo de las características y competencias ineludibles para el cargo.

² Instituto: formación inicial que reciben los participantes de Enseña por Colombia



Según Atcom (2018) un tipo de selección de personal es el Interno “este tipo de reclutamiento trata de la necesidad de cubrir un puesto de trabajo generando promociones dentro de la misma organización, es decir se brinda la oportunidad a un trabajador de la empresa para que emprenda nuevos desafíos”. Este tipo de reclutamiento se da mediante “promociones, transferencias, programas, ascensos, plan carrera” (Renteria, 2008). Resulta más cómodo y sencillo en cuanto la persona que asciende o cambia de puesto ya conoce las políticas de ese entorno laboral. Este tipo de selección resulta “más económico ya que la empresa no invierte en anuncios, además es rápido, valido y seguro” (Renteria, 2008).

Por otra parte, está el tipo de selección Externo que es “cuando se presenta la necesidad de cubrir un puesto de trabajo específico que solamente debe buscarse fuera de la compañía por varias razones tales como; profesión específica, tareas técnicas específicas que solamente las puede cubrir tal especialista”, (Atcom, et. al. 2018). La contratación o selección de este tipo trae beneficios en cuanto una nueva persona en la empresa o institución puede traer nuevos aprendizajes y propuestas innovadoras, pero para atraer buenos empleados, es importante como lo expone López (2010) saber que “un anuncio debe ser directo y claro y constituye el reflejo de la organización y la imagen que se desea comunicar”.

Se encuentra también el tipo de selección Mixto el cual, según Atcom (2018) “este tipo de reclutamiento personal toma en cuenta la opción de promover a un trabajador interno, como reclutar a un trabajador desconocido. Esta consideración se basa generalmente en donde se miden los talentos versus la remuneración”. Este tipo de selección puede decirse que implica más costo en la contratación porque se implementan ambas metodologías de reclutamiento, pero puede ser muy eficiente a la



hora de buscar una persona acorde para un cargo supremamente importante en la organización.

Por último, se encuentra el tipo de selección Reclutamiento 2.0 este tipo de reclutamiento “hace alusión a la forma de reclutar profesionales que comparten redes de negocios y talentos en donde son ellos mismos los que se publicitan y promueven sus experiencias y motivaciones profesionales” (Atcom, 2018). Este tipo de selección es nuevo y está relacionado con el avance de las redes sociales, resulta práctico y versátil, y un ejemplo de este tipo es la aplicación LinkedIn, donde empresa y personas naturales preparan un perfil profesional para encontrar trabajo o empleados.

Es importante resaltar que en cuanto al proceso de selección también puede ser interno o externo, como lo expone Castaño, López y Prieto,(2011) “el encargado de llevar a cabo el proceso de reclutamiento y/o el proceso de selección puede ser interno (generalmente departamento de Recursos Humanos) o externo a la empresa demandante”, esto pasa cuando las empresas contratan temporales que les ayudan a encontrar los empleados que necesitan, más que todo cuando son contrataciones masivas. Según un estudio de Naranjo (2017) “se evidenció que el 58% del personal prefiere el reclutamiento externo porque considera que es mayor la probabilidad de ser vinculado a la empresa” (p.91).

Proceso de selección docente

La labor docente es una de las más importantes en la sociedad, ya que de estos profesionales depende la formación de cientos de ciudadanos, y por ello, se debe procurar un buen proceso de selección que conlleve a tener los mejores formadores dentro de las instituciones educativas sin importar su modalidad. Lógicamente cada institución ya sea pública, privada, de educación primaria, secundaria o universitaria tiene sus propias políticas de selección, pero por lo general estas son similares y están



basadas en las competencias. Según (Gouveia & Montiel, 2006) “la Organización Internacional del Trabajo [OIT] (2000), ha definido a las competencias laborales, como el conjunto de saberes (saber, saber hacer, saber estar, y saber ser- conocimientos, procedimiento y actitudes) combinados, coordinados e integrados en el ejercicio profesional”.

A continuación, se menciona un ejemplo del proceso de selección de una Institución, procesos con los cuales se busca evaluar de manera integral las competencias que se deben tener para ejercer la labor docente. El primero es la “Redacción del ensayo: Mediante la redacción del ensayo se evaluarán las destrezas de composición y análisis en las áreas de contenido, desarrollo del tema, organización, cohesión, gramática y análisis” (Universidad del Este, 2015). Es supremamente importante que un docente tenga habilidad de escritura, de síntesis, interpretación y reflexión, ya que esto le da capacidad de abordar proyectos o de tener introspección desde sus encuentros pedagógicos.

Se encuentra la Entrevista grupal “(total de 20 minutos), en la actividad se realizan preguntas relacionadas con la hoja de vida o currículo de cada candidato; así como otras preguntas realizadas por los evaluadores que tendrán como propósito valorizar las destrezas de relaciones interpersonales y de comunicación” (Universidad del Este, 2015), esto resulta conveniente porque el diálogo entre docentes permite también identificar cómo será su relación con los compañeros y comparar ideas innovadoras entre los mismos.

Por último, la Clase demostrativa (20 minutos por candidato) donde “el candidato deberá demostrar sus competencias como docente y facilitador del aprendizaje. Los criterios a ser evaluados son los siguientes: el contenido del tema y organización,



ortografía, habilidad de comunicación, estimulación del pensamiento crítico, entusiasmo y dominio del tema”. Este modo de evaluación para el proceso de selección es uno de los más certeros porque indica la formación pedagógica, didáctica, el manejo de grupo y la manera en que se abordan las temáticas dependiendo del grado escolar que se le asigne.

La Institución Educativa Gutiérrez Cabrera (2015) resalta otros aspectos a tener en cuenta en el proceso de selección docente, el dominio de la disciplina, es decir, que sepa de su área de conocimiento, que pueda abordar todos los objetivos de la malla curricular. Las características personales, que tenga buenos valores que le permitan buenas relaciones con sus colegas, directivos, padres de familias y estudiantes, así mismo las competencias docentes, entendiéndolas desde las capacidades relacionadas con el modo de llegar a sus estudiantes en sus clases y por último, debido al avance tecnológico, se busca competencias tecnológicas en el docente, que le permitan comprender la importancia del uso de las TIC para los estudiantes de esta nueva era, además de que pueda manejar las herramientas tecnológicas e innovar.

Hablando de proceso de selección a nivel de Colombia, se puede rescatar la metodología del concurso docente, el cual se realiza para instituciones públicas y donde el docente se vincula al Estado para hacerse cargo de una plaza a nivel rural o urbano. Para el concurso el docente debe tener un grado de profesionalidad y concurra dependiendo de este, el primer filtro es a través de un examen de competencias lógico matemática, interpretativas y pedagógicas. Los docentes que ganen el concurso con un puntaje determinado presentan papeles y entrevistas, y al momento de llegar a la institución entran en un periodo de prueba. Para algunos tipos de ascenso en esta



modalidad el docente debe presentar un video de sus clases (Ministerio de Educación Nacional, 2020).

Existen diferencias en el proceso de selección e instituciones de educación pública o privada, por lo general las públicas se rigen en el sistema gubernamental que maneja concurso, provisionales y banco de hojas de vida, el privado por su parte puede optar por convocatorias generales y su proceso de selección o los ítems a evaluar dependerán de las políticas de la institución, pero siempre en cada proceso se busca evaluar las competencias de manera integral.

METODOLOGÍA

Machine learning es la metodología que se utiliza en esta investigación para establecer un modelo que nos permita tomar decisiones de selección. Mitchell (1997) provee la siguiente definición de machine learning “Se dice que un programa de computadora aprende de la experiencia E con respecto a alguna tarea T y mide su desempeño P, si el desempeño en la tarea T, es medida por P, mejora con la experiencia E. ". En este estudio la variable E es una base de datos de entrevistas con clasificaciones determinadas. T clasificar Alumnis y P porcentaje de individuos que terminan el programa.

El machine learning tiene dos tipos de clasificación, (1) Supervised learning, conocido en español como aprendizaje supervisado donde el investigador ya conocer la variable de respuesta, y (2) Unsupervised learning, aprendizaje no supervisado cuando se tienen una serie de datos y se quiere ver la asociación natural entre ellos. En este caso se toma

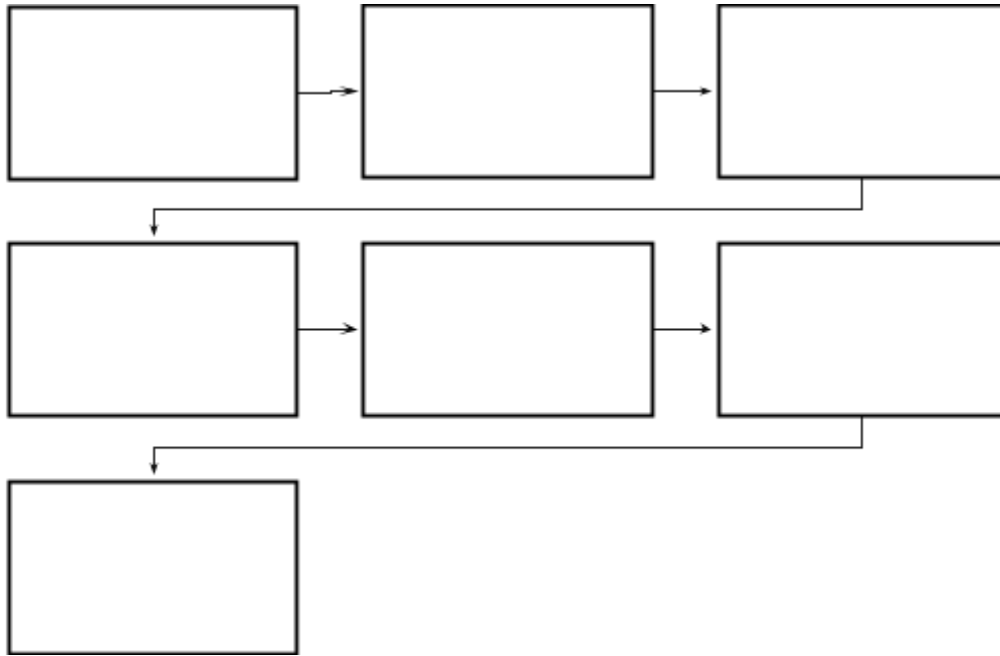


la decisión de hacer un análisis supervisado pues se conoce quienes son los individuos que terminan el proceso.

Dentro del aprendizaje supervisado existen dos tipificaciones (1) de regresión cuando la variable de respuesta es continua, es decir, datos que pueden tener cualquier valor entre un intervalo y (2) de clasificación para variables discretas o datos que no toman valores entre consecutivos. En este caso la variable de respuesta hace referencia a individuos que terminan el programa de Enseña por Colombia, por lo que es conveniente seleccionar un modelo de machine learning supervisado y de clasificación.

Los bosques aleatorios o Random forest es un algoritmo popular y muy eficiente para problemas de clasificación y regresión (Breiman, 2001). Pertenece a la familia de métodos conjuntos, que apareció en el aprendizaje automático a finales de los noventa. El principio de los bosques aleatorios es combinar muchos árboles de decisión contruidos usando varias muestras bootstrap provenientes de la prueba de entrenamiento y eligiendo aleatoriamente en cada nodo un subconjunto de variables explicativas X .

La metodología usada se explica con el siguiente diagrama:



1. Recolección de los datos: Es muy importante entender cómo se recolecta y se procesa la información en la organización. Por esta razón, se profundiza en el candidate journey de Selección de Enseña por Colombia donde se evidencian los siguientes momentos claves y se recibe la información que copila las competencias aquí mencionadas y posteriormente serán analizadas:

- La primera fase de selección que tienen que atravesar los profesionales que quieran ser parte de esta experiencia es la parte de registro y postulación, en esta fase Enseña por Colombia reconoce la formación académica, profesional y laboral, así como los conocimientos que se tienen sobre la organización. En esta fase se almacenan aproximadamente 35 variables entre cuantitativas y cualitativas entregadas por los candidatos que son parte de la formula.
- La segunda fase es una serie de pruebas virtuales: donde se analizan las competencias de locus de control, pensamiento crítico y Mindset. En esta segunda fase se almacenan 50 variables cualitativas.



- Posterior a la evaluación, se realiza una fase de entrevista con los seleccionados donde se exploran los atributos esenciales que debe tener todo participante para que pueda sumar con la misión y visión de la organización. La última fase con 31 variables cualitativas.
2. Preparación de los datos:
- a. Análisis exploratorio:
- La primera base de datos analizada contaba con 2744 individuos y 171 variables sin incluir la de respuesta desde el 2014 al 2018.
 - Densidad de datos: revisar la densidad de datos que presentaba cada individuo. Se encontró que la cohorte 2014 presentaba la mayor cantidad de datos ausentes, al no contar con la medición de todas variables que si se midieron en los años posteriores. Esto se debe a que en el 2015 se incorporaron nuevas variables, por lo tanto, se procede a analizar las cohortes del 2015 al 2018.

	2014	2015	2016	2017	2018
% de datos	29,34 %	62,90%	57,37%	61,22%	69,76%
Numero de variables	74-93	128	128	128	128

- b. Ajuste de los datos:
- Datos faltantes: la revisión de datos faltantes es muy importante pues el modelo no permite que existan. Hay individuos que no se presentaron al día de entrevista por lo que no tenían su proceso completo, aunque si habían realizado la primera revisión y la actividad virtual, razón por la cual se procede a eliminar dichos registros; se busca analizar individuos que hayan realizado todo el proceso de



selección. Luego, se evidenció individuos que por llegar tarde a algunas de las actividades no tenían información en las competencias de la actividad, por lo tanto, se procede a colocar el mínimo valor posible en la variable. Y finalmente, para los individuos que les faltaba un solo dato en las variables por competencias, se procede a una imputación de datos con el valor más probable.

- Un único individuo por toda la base de datos: se necesita que solo un individuo se represente en toda la base de datos. En las personas que se habían presentado en día de entrevista en diferentes años, se seleccionó el dato donde el candidato tuvo el mejor performance.
 - Variable de respuesta: se realiza la búsqueda de los individuos que terminaron el programa en esas fechas y se procede a la categorización.
- c. Decisiones finales: al notar la cantidad de datos almacenados y las limitaciones del algoritmo Random Forest, se procede a seleccionar únicamente la parte del proceso que compone la entrevista con 31 variables.
3. Selección del modelo: se selecciona el modelo random forest pues nos permite hacer una gran proyección con varios árboles de decisión.
4. Entrenar el modelo
- a. Aleatorizar la base de datos: con el fin de poder tener un mejor ajuste del modelo.
- b. Dividir los datos: se agrupan en datos de entrenamiento y de testeo, en consecuencia, se genera el modelo con la primera y con la segunda se prueba.
- c. Crear modelo: con la data dividida se establece el modelo random forest con los datos de entrenamiento.



5. Se evalúa el modelo

Se evalúa el modelo con la base de datos de entrenamiento y luego la de testeo. De los 1699 el 11.39% era Alumni, los resultados arrojaron un gran ajuste, pero para predecir las personas que no pasaban el modelo. Para lograr el objetivo de generar un modelo de predicción de las personas que terminaran el modelo, se reorganiza y filtra la base de datos con la información recolectada de personas que fueron admitidas en Enseña por Colombia³. La nueva data tiene 325 individuos donde el 58% son Alumnis y 42% Individuos que fueron seleccionados pero que no culminaron el proceso de formación.

RESULTADOS

1. Modelo Inicial: Esta sección describe la simulación inicial con el algoritmo random forest. En el programa R se incorpora los paquetes randomForest y Caret para generar el modelo. Se procede a analizar un nuevo modelo con los 325 individuos. El 70% se ubica en la base de datos entrenamiento y el 30% en la de testeo. La exactitud de una predicción de un Random forest se puede estimar a partir de estos OOB (el error del modelo). El primer modelo nos predice clasificar a cada uno de los individuos con un 55% de confianza, y un OOB de 45%. Para mejorar el modelo se debe ajustar mejor los parámetros.

```
Modelo <- randomForest(x = pfl[training_mod, 2:32], y =  
pfl[training_mod, 1])  
  
Type of random forest: classification  
Number of trees: 500  
No. of variables tried at each split: 5
```

³ La selección de los individuos que fueron admitidos en el proceso se basa en los datos almacenados en su plataforma de selección.



LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

Reference

Prediction	Alumni	Registro
Alumni	44	31
Registro	12	10

Accuracy : 0.5567

95% CI : (0.4523, 0.6576)

No Information Rate : 0.5773

P-Value [Acc > NIR] : 0.697577

Kappa : 0.0316

McNemar's Test P-Value : 0.006052

Sensitivity : 0.7857

Specificity : 0.2439

Pos Pred Value : 0.5867

Neg Pred Value : 0.4545

Prevalence : 0.5773

Detection Rate : 0.4536

Detection Prevalence : 0.7732

Balanced Accuracy : 0.5148

'Positive' Class : Alumni

2. Ajuste de parámetros: se examina si la tasa de error OOB se ve afectada sustancialmente por cambios en mtry, ntree.

a. Se busca el mejor mtry (número de variables aleatorias como candidatas en cada ramificación). Kuhn y Johnson (2013, como se citó en Breiman (2001)) recomendó $mtry = \sqrt{p}$, siendo p el número de predictores para bosques aleatorios de clasificación e indicó que mtry p debería mejorar el rendimiento predictivo del modelo. Un mejor mtry incrementa la precisión (accuracy) de la predicción. Recientemente, Genuer, Poggi y Tuleau (2008) presentaron resultados de simulación que indican una mejor precisión de predicción con $mtry = p$ para los datos artificiales del artículo de Friedman (1991). Dentro del estudio de simulación, el mtry que minimizó OOB fue de 2. La técnica Cross Validation nos ayudó con esta predicción, de forma que evalúa el modelo repitiendo y calculando la media aritmética en diferentes partes de la base de datos. En este caso se



divide la data en 10 partes. Esta técnica es muy utilizada para validar modelos generados por inteligencia artificial como el random forest.

```
, 326 samples
31 predictor
2 classes: 'Alumni', 'Registro'
```

No pre-processing

Resampling: Cross-Validated (10 fold)

Summary of sample sizes: 293, 293, 294, 293, 294, 294, ...

Resampling results across tuning parameters:

mtry	Accuracy	Kappa
2	0.5706439	-0.01818612
45	0.5006629	-0.08966238
89	0.4884470	-0.10371052

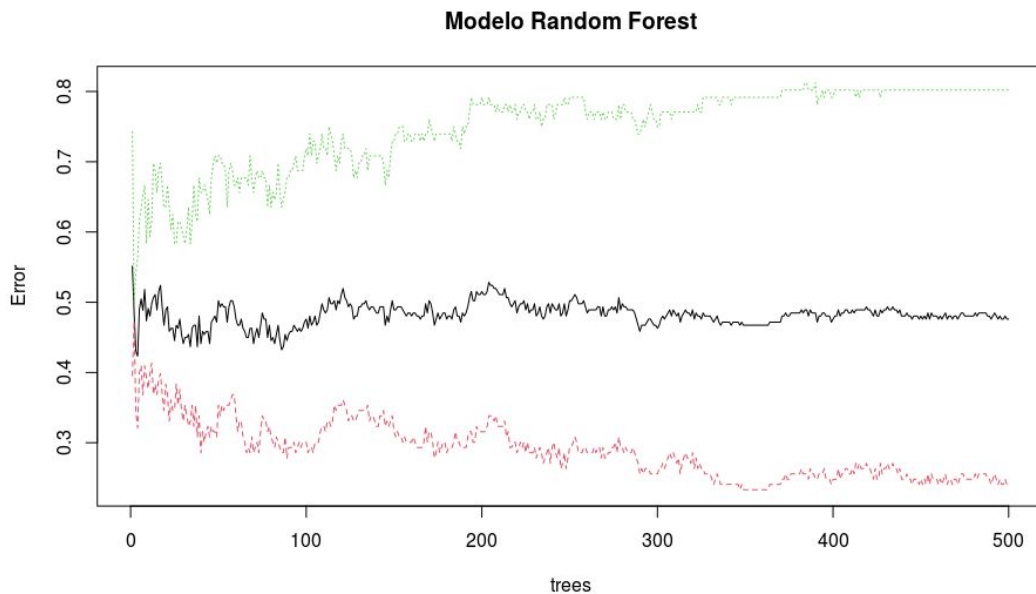
Accuracy was used to select the optimal model using the largest value.

The final value used for the model was mtry = 2.

b. Se busca el mejor `n`tree (número de árboles en el bosque) para estabilizar el error, pero usar demasiados árboles puede ser innecesario e ineficiente. Un Random Forest por lo general consta de un gran número de árboles, por ejemplo un `n`tree de 1000. Los Random Forestes son aleatorios de dos maneras: (1) cada árbol se basa en un subconjunto aleatorio de las observaciones, y (2) cada división dentro de cada árbol se crea con base a un subconjunto aleatorio de variables candidatas `mtry`. Los árboles son bastante inestables, por lo que esta aleatoriedad crea diferencias en las predicciones de los árboles individuales. La predicción general del bosque es el promedio de las predicciones de los árboles individuales; debido a que los árboles individuales producen funciones escalonadas multidimensionales, su promedio es nuevamente una función escalonada multidimensional que, no obstante, puede predecir funciones suaves porque agrega una gran cantidad de árboles diferentes. Lo propuesto por Friedman (1991) para la visualización de los resultados del bosque, se pueden utilizar gráficos de efectos principales y de interacción. Dentro del estudio de simulación se graficó el modelo



donde se compara el error con el número de árboles. Se selecciona un ntree de 500, debido a que es el punto en el cual el error tiende a tomar un valor estable.



3. Modelo ajustado: Se incorporan los parámetros encontrando mtry de 2 y un numero de 500 árboles

```
Mod.1<-randomForest(x=pfl [training_mod,2:32],  
+                    y=pfl [training_mod,1],  
+                    ntree = 500,  
+                    mtry = 2,  
+                    importance = TRUE)
```

Confusion Matrix and Statistics

	Reference	
Prediction	Alumni	Registro
Alumni	47	31
Registro	9	10

Accuracy : 0.5876
95% CI : (0.4831, 0.6867)
No Information Rate : 0.5773
P-Value [Acc > NIR] : 0.4611994

Kappa : 0.0896

Mcnemar's Test P-Value : 0.0008989

Sensitivity : 0.8393
Specificity : 0.2439



LOS LIBERTADORES

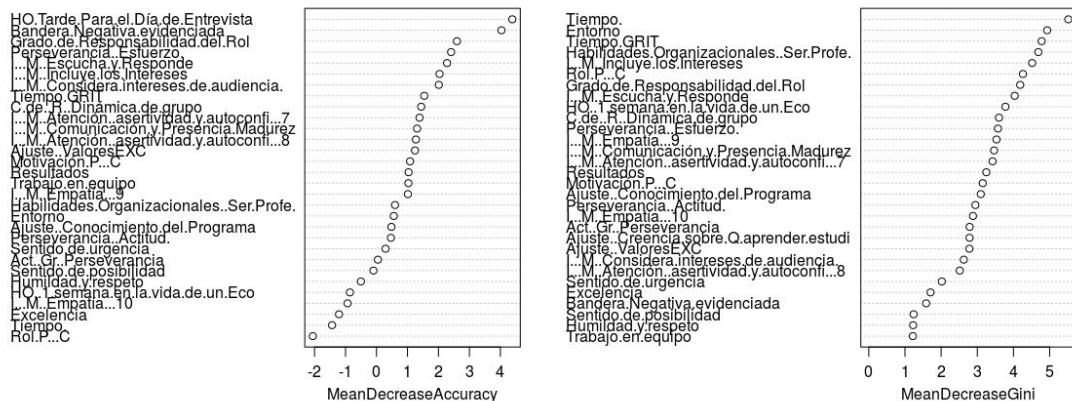
FUNDACIÓN UNIVERSITARIA

Pos Pred Value : 0.6026
Neg Pred Value : 0.5263
Prevalence : 0.5773
Detection Rate : 0.4845
Detection Prevalence : 0.8041
Balanced Accuracy : 0.5416

'Positive' Class : Alumni

4. Las variables que más le aportan al modelo: el denominador de la variable importancia es la cantidad total de variación de respuesta explicada por todos los k componentes. Por lo tanto, cuanto mayor sea el peso normalizado y la cantidad de variación de respuesta explicada por el componente, más importante será el predictor en el modelo.

Importancia modelo final



	Alumni	Registro
Habilidades.Organizacionales..Ser.Profe.	0.0020625601	-0.0012971302
I...M..Incluye.los.Intereses	0.0024847491	0.0019114213
I...M..Comunicación.y.Presencia.Madurez	0.0012583806	0.0014942531
I...M..Escucha.y.Responde	0.0030868250	0.0019718988
C.de..R..Dinámica.de.grupo	0.0010524594	0.0021515689
I...M..Atención..asertividad.y.autoconfi...7	0.0011697196	0.0018907545
I...M..Atención..asertividad.y.autoconfi...8	0.0010641190	0.0013472769
I...M..Empatía...9	0.0017489711	0.0005114754
I...M..Empatía...10	0.0002282156	-0.0025039061



Act..Gr..Perseverancia	0.0016787795	-0.0019656227
Grado.de.Responsabilidad.del.Rol	0.0044432961	0.0018486618
Resultados	0.0043924491	-0.0032495037
Perseverancia..Esfuerzo.	0.0041947743	0.0011032032
Perseverancia..Actitud.	0.0002304630	0.0006767009
Rol.P...C	-0.0020820341	-0.0024776220
Tiempo.GRIT	0.0003726611	0.0039297324
Entorno	-0.0010162477	0.0029007545
Motivación.P...C	0.0008956439	0.0008910100
I...M...Considera.intereses.de.audiencia.	0.0019109695	0.0015470369
Tiempo.	-0.0039614697	0.0015979150
Bandera.Negativa.evidenciada	0.0022885236	0.0028542035
Ajuste..Creencia.sobre.Q.aprender.estudi	-0.0027991097	-0.0056350873
Ajuste..Conocimiento.del.Programa	0.0025057561	-0.0018245740
Ajuste..Valores Exc	0.0021932093	-0.0007185977
Sentido.de.posibilidad	0.0001049005	-0.0002329599
Perseverancia	-0.0013696592	-0.0004327010
Humildad.y.respeto	-0.0003799782	-0.0001951148
Trabajo.en.equipo	0.0012374371	-0.0003177712
Sentido.de.urgencia	-0.0003907865	0.0013102732
H0..1.semana.en.la.vida.de.un.Eco	-0.0015773559	0.0001267884
H0.Tarde.Para.el.Día.de.Entrevista	0.0017882057	0.0011494008

CONCLUSIONES

El objetivo de la investigación es proveer a Enseña por Colombia con un modelo predictivo de selección de profesionales que terminen los dos años de programa. La metodología random forest es un buen método capaz de resolver problemas de naturaleza regresiva y de clasificación debido a que las variables que se consideran para resolver el problema dan una amplia gama de posibles resultados, aumentando la capacidad de obtener una predicción precisa. Herramientas como R, permiten obtener mejores resultados gracias a su gran capacidad computacional para almacenar y procesar grandes flujos de información.

En el proceso para dar el mejor ajuste se realizaron diversas iteraciones con el número de árboles para obtener el accuracy más preciso. Este factor fue determinante para obtener una clasificación con un 58% de exactitud de los candidatos seleccionados.



Es indispensable para Enseña por Colombia seguir buscando mejorar el proceso de selección, para lograr impactar en la educación con altos estándares educativos. La mejora en este proceso puede aportar y afianzar a la relaciones y credibilidad que tienen ExC con sus organizaciones aliadas y donantes.

RECOMENDACIONES

Para futuros análisis relacionados con el proceso de selección de Enseña por Colombia se sugiere revistar los dos trabajos de regresión realizados por Diego Sánchez y de Santiago matamoros. Por otro lado, para los públicos que estén interesados en profundizar el análisis por medio del Machine Learning se sugiere considerar modelos para abordar esta problemática con el algoritmo K-NN que resuelve problemas de clasificación y de regresión y Redes neuronales.

REFERENCIAS BIBLIOGRÁFICAS

Atcom. (17 de septiembre de 2018). Atcom. Obtenido de Atcom:
<https://www.atcom.cl/4-tipos-de-reclutamiento-del-personal.php>

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification

And Regression Trees, Chapman & Hall, New York

Breiman, L., 2001. Random Forests. Machine Learning. 45, 5-32.

Burger 2018. Introduction to Machine Learning with R. Tree-Based Methods 135-156

Cutler, C. Edwards Jr (2017). Random forests for classification in ecology,

Castaño, M., Lopez, G., & Prieto, J. (2011). Guía Técnica y de buenas prácticas en reclutamiento y selección de personal . Colegio Oficial de Psicólogos de Madrid , 46.



Enseña por Colombia(2020) <https://www.ensenaporcolombia.org/>

Genuer, Poggi, Tuleau-Mal (2010). Variable Selection using Random Forests

Gouveia, E., & Montiel, K. (2006). El proceso de selección del personal docente basado en el enfoque de competencias. Una propuesta. Telos , 199 - 214.

Institución Educativa Gutiérrez Cabrera. (2015). Procesos de selección docente en un modelo educativo basado en competencias. secretaria académica.

Kuhn, Johnson (2013) Applied Predictive Modeling. Regression Trees and Rule-Based Models 173-218

Lopez, R. (2010). La selección de personal basada en competencias y su relación con la eficacia organizacional. Perspectivas, 129 - 152.

Mitchell (1997) Machine Learning. Introduction, 3.

Ministerio de Educación Nacional. (2019). Resolución 016720 del 27 de diciembre de 2019. Congreso de la República.

Ministerio de Educación Nacional. (2020). Ministerio de Educación Nacional. Obtenido de Ministerio de Educación Nacional: https://www.mineduacion.gov.co/1759/w3-propertyvalue-48460.html?_noredirect=1#:~:text=El%20concurso%20de%20m%C3%A9ritos%20para,determina%20su%20inclusi%C3%B3n%20en%20el

Naranjo, R. (2017). El proceso de selección y contratación del personal en las medianas empresas de la ciudad de Barranquilla (Colombia). pensamiento y gestión, N° 32, 32.



LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

Renteria, M. (2008). Proceso de reclutamiento y selección de personal docente de la
Strobl (2007).Bias in random forest variable importance measures: Illustrations, sources
and a solution.

Universidad Pedagógica Nacional . Universidad Pedagógica Nacional.

Universidad del Este. (2015). Normas y Procedimientos para la Selección del
profesorado. Universidad del Este.

Teach for all (2020) <https://teachforall.org/>